



Věc: **Zpráva z pracovní cesty**  
Pracovník: **Pavlaína Rouhová**  
Akce: **ProArc - jednání**  
Datum konání: **3. června 2016**  
Místo konání: **Oddělení digitalizace, Městská knihovna Praha**

#### ProArc (obecně):

- je systém pro produkci digitálních dokumentů a jejich archivaci, umožňuje rychlou částečně automatizovanou produkci digitálních dokumentů zahrnující popisná, strukturální a archivační metadata i OCR
- koordinátorem vývoje je Knihovna AV ČR (M. Lhoták)
- na Githubu jsou stránky ProArcu – informace, dokumentace, novinky, FAQ, komunikace, diskuse, problémy: <https://github.com/proarc/proarc/wiki>, až budeme uživateli ProArcu, je třeba se sem zaregistrovat (issues)
- na Googlu je též mailová diskusní skupina ProArcu: <https://groups.google.com/forum/#!forum/proarc-users> (registrace)
- open source, Fedora Commons repository (Fedora repozitář), Java, PostgreSQL
- je kompatibilní se systémem Kramerius a podporuje standardy NKP pro digitalizaci (plná podpora MODS, příprava podpory ADM a TECH metadat): <http://www.ndk.cz/standardy-digitalizace/metadata>
- využívá standardy pro dlouhodobou archivaci
- je implementován OAIS model a archivační ISO standardy
- umožňuje editaci metadat a vyhledávání v metadatech
- podporuje dávkové/hromadné úpravy
- umožňuje vytváření zcela nových objektů
- podporuje přebírání dat z externích systémů (Aleph/RDCZ)
- Fedora objekt integruje metadata a digitální obsah (sken, ...)
- generuje UUID
- podporuje přidělování URN:NBN
- export PSP balíčků nebo XML
- umožňuje rozšíření o dalších datové modely

#### MKP

- <https://www.mlp.cz>
- v MKP – vlastní knihovní (archaický) systém Koniáš, chtěli by přejít na nový knihovní systém, ale je to problém, mají mnoho poboček (42?) – bylo by to moc nákladné
- nadstavbou Koniáše je Marcipán (Marc...)☺ – systém pro implementaci MARC21 a RDA
- jména serverů knihovny jsou podle zvířat, hlavní server je Orel, vedlejší Sumec a Jeseter ☺
- MKP má nejstarší hudební knihovnu u nás

#### Oddělení digitalizace MKP

- <https://www.mlp.cz/cz/kontakty/pobocky/oddeleni-digitalizace>
- v současnosti 7 a půl úvazku, z projektu propustili 9 lidí - projekt byl na jeden rok, všichni dělali všechno – skenování, ořez, OCR, metadata – mimo periodik, ty dělají pouze tři lidé



- nově datové úložiště se serverem – dvě datová pole, každé o velikosti 18TB hrubé kapacity - jedno pole je v oddělení digitalizace v Provaznické ulici a druhé v Ústřední knihovně na Mariánském náměstí, na pole v Ústřední knihovně probíhá záloha dat z úložiště v oddělení digitalizace
- tři (čtyři) knižní skenery – jedná se o skenery CopiBook Cobalt HD (pro snímání klasických formátů dokumentů) a DigiBook Suprascan II (pro snímání dokumentů nestandardních formátů) BookEye a Zeutschel (záložní)
- výkonné počítačové stanice, monitory vhodné pro grafické zpracování obrazu včetně krytů proti bočnímu světlu
- skenery skenují barevně do formátu TIFF (ne JPEG 2000) až do rozlišení 600 dpi
- software pro dávkové zpracování obrazu - software LIMB od firmy i2S
- OCR software ABBYY Recognition Server 4 s licencí na zpracování 30 000 stran za měsíc
- dříve používali Sirius, nyní ProArc
- pro složení (spojení) velkých map rozdělených při skenování používají Adobe Photoshop
- ke skenování používají program IrfanView
- na ořez používají software LIMB
- všechny původní (originální) skeny si zálohují
- ořezávají vnějškem
- OCR - textů ve švabachu mají málo – na to mají Frakturu s malou licencí, když toho mají víc, nechají si to udělat externě (<http://www.frakturschrift.com/en:start>)

#### Digitalizace - ProArc (MKP)

- <http://proarc.mlp.cz/proarc/>
- ProArc pracuje v AACR2 (v retrodigitalizaci je to povoleno), když stáhnou data v jiných polích (RDA), tak je přepíše do alternativních polí
- ProArc - chystá se přechod na RDA v rámci aktualizace (po prázdninách?)
- uživatelský návod pro ProArc bude – je v přípravě (po prázdninách?)
- složky:
  - složky dokumentů pojmenované signaturou dokumentu
  - vyhledávání podle signatury
  - vytvořit složky Import a Export – vždy po použití je vymazat
  - vytvořit složku Doplněk – pro doplňkové dokumenty
- v nové verzi ProArcu bude vpisovatelné XML
- výsledky exportují rovnou do NDK
- lze pracovat s prázdnými stránkami
- lze indexovat jako v Total Commanderu
- některé funkce je třeba opětovně nahodit (aktualizovat), přepnout do jiného modulu a pak zase zpět
- URN:NBN (Czidlo) – tlačítko v ProArcu – automatické propojení do NKP (NDK) – velmi rychlé
- pořád ukládat (během práce)!
- Ukládá do repozitáře Fedora (součást ProArcu)
- je domluvena spolupráce – když bude třeba, mohu se na ně kdykoliv obrátit
- je předem (na neurčito) domluvena další schůzka – patrně až už budeme mít ProArc a nějaké konkrétní otázky
- Ondra má domluvenu spolupráci s IT pracovníkem
- máme OAIS? Ondra říká, že jo???